



Assignment 1:

Memory-based CF for Rating Prediction

Overview

- Programming
 - UCF or ICF for rating prediction
- Requirements
 - Project report
 - Executable source code
 - Prediction results

Project Report

- Main content
 - Detailed implementation introduction
 - Key code explanation
 - Result analysis
 - Hyper-parameter influence
 - Table or plot
 - How to run your code
 - Environments (E.g., python 2.X or python 3.x, c++?)
 - Dependency libraries

Source Code

- Programming language
 - Python is preferred.
- Terminal command
 - E.g., `python2.7 xxx.py test.file output.file`
- Source code files
 - How the method is implemented

Data Format

- Train set
 - With rating scores

```
train.csv
1 user_id,business_id,date,stars
2 A2JGzkvNjckSmps_4FbKWw,Xg5qEQiB-7L6kGJ5F4K3bQ,2014-03-18 01:14:10,5.0
3 rypcWiSNGM0suWsiSLh9xA,4RoTEeqB_MNn6yaqZmlZHg,2015-08-29 18:32:15,4.0
4 Dgk0Wdoh7HPjhKQEPBU_jQ,ZOmF-3NN4Z59b2Fw6VAM7g,2015-09-14 16:33:03,3.0
5 FIk4lQQu1eTe2EpzQ4xhBA,HK2Ki-PvnNN-YMTlX1uSVA,2012-09-29 02:03:42,4.0
6 VizhcyMWWPz3UDXEBeix4w,UPIYuRaZvknINod1w8kqRQ,2011-06-10 20:35:42,3.0
7 2EuPAGalYnP7eSxPgFCNDg,E83nSU_y9zed0zQnkTjV1g,2017-08-14 18:56:00,2.0
8 WXlxViTwXHPBvhioljN9PQ,IRzY7yoBqoHaZNN08WiWQQ,2016-04-30 14:42:50,4.0
9 cMeTAiW60I5wE_vLfTxoJQ,DESv2ys6SjBKA4SyDtJvxw,2012-06-26 00:44:03,4.0
```

- Test set
 - Without rating scores

```
test.csv
1 user_id,business_id,date
2 PfpRvMAESbC2bC8FUIMdNg,Kbbm6Vd5UdbP10dwjBghRw,2018/10/15 0:52
3 oaaEXgQ3x51cXE3GTxrT1Q,2GmGT-7QjowR1ihup3FbVA,2011/11/27 9:12
4 yT_QCcnq-QGipWWuzIpvtw,pOEL97ld-FJMK08Ki8JmYg,2016/3/11 19:09
5 fRVNHA12RjosC67Y67G3cA,UklWme3kwg6L9rd4tCNB15w,2016/9/11 15:53
6 48vRThjhuhisQINQ2KV8Sw,LNGBEEelQx4zbfWnlc66cw,2011/3/23 3:22
7 q5FQmuXxzPEsvEtA_Mvd1w,fSBhe0A6Dfa8JCYccfpMog,2013/8/9 22:06
8 W0VE9M7Dikrpol8j1_QqyQ,3oTVApC-eUzpGjrOVxIr5g,2017/9/7 2:27
9 avmRUkWovTsaDqKiNKdivQ,wUKzaS1MHg94RGM6z8u9mw,2012/10/25 19:31
```

Prediction Results

- A file containing prediction results
 - Format: each line corresponds to one prediction tuple
 - User_id, Business_id, Rating_score
 - E.g., A2JGzkvNjckSmps_4FbKWw, Xg5qEQiB-7L6kGJ5F4K3bQ, 4.0

Submission

- File name
 - zip file, named with 'Name+ID+Assignment No.'
- Submission URL
 - <http://xzc.cn/Ne34O6sqWw>

第一次作业

第一次作业提交

文件提交区域

* 提交人姓名:

→ 拖拽文件上传
(或点击)

最多上传20个文件，单个文件最大20M

提交

How to evaluate?

- Code
 - Executable
 - Clear, easy-to-understood
- Prediction results
 - Effectiveness
 - How is the predicted score compared with ground-truth?
 - RMSE
- Report
 - Integrity
 - Readability
 - Highlights

Report Example

数据挖掘第一次编程作业

——实现数据挖掘 Apriori 算法

2018 年 3 月 26 日

目录

1 算法简介	2
1.1 简介	2
1.2 伪代码描述	2
1.3 实现过程描述	3
2 核心代码注解	3
2.1 1-项集的产生	3
2.2 计时	3
2.3 Apriori_gen 过程	4
3 结果分析	4
3.1 频繁集发掘	4
3.2 关联规则发掘	4
4 代码运行	4
4.1 环境	4
4.2 依赖	6
4.3 命令行命令	6
4.4 Apriori.py	6
4.5 asso_gen.py	6
5 提交文件列表	7
5.1 Apriori.py	7
5.2 asso_gen.py	7
5.3 out 目录	7
5.3.1 频繁集文件 out.dat	7
5.3.2 关联规则文件 association_rules.dat	7
5.3.3 关联规则二进制中间文件 fdict.pkl	7

2 核心代码注解

以上的 Apriori 算法过程易于理解,但是实现时还是存在一些 python 语言层面的问题,比如,数据结构的选择,以及相应的操作选择。故以下对核心代码做必要说明,余下部分不做过多赘述。

2.1 1-项集的产生

对于算法中项集-频数形式的 k-v 对,在 python 中采用 dict 存储,但是 dict 中 key 值不能为 set 和 list 等无法哈希的数据类型,故此处采用 frozenset 这一类集合结构。

```
1 @timer
2 def generate_1_items_dict(translit):
3     """ insert and count 1-items occurrences in a hash tree """
4     l1 = {}
5     for trans in translit:
6         for item in trans:
```

不同支持度下各项集元素个数

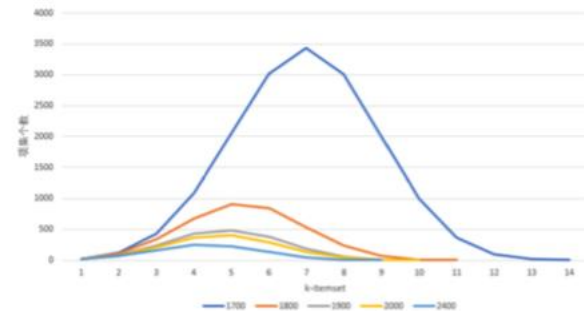


图 2: 不同支持度下各项集元素个数变化。随着最小支持度的变大,各项集元素个数也总体下降,而当最小支持度较小时,项集元素个数会非常大,例如上图中最高点达到近 3500 项。

Data and Deadline

- **Kaggle url**

- <https://www.kaggle.com/t/450e7cd913014ec28a261995d8c7b3d9>

- **Deadline**

- 24:00, 2022-4-3
- Submission files after that time might be penalized

← → ↻ kaggle.com/c/ecnuRSclass/overview

应用 YouTube 地图 Gmail 科研 学生指导

≡ kaggle

+ Create

🏠 Home

🏆 Competitions

📁 Datasets

<> Code

💬 Discussions

🎓 Courses

▼ More

📁 Your Work

RECENTLY VIEWED

👤 助教的示例代码sample_...

📁 ECNU recommender s...

🏠 View Active Events

🔍 Search

Community Prediction Competition

ECNU recommender systems class

This is a competition of RS class for undergraduates in ECNU.

14 days to go

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#)

[Join Competition](#) ...

Overview

[Description](#)

ECNU推荐系统本科课程作业提交

这里是张伟老师开设的推荐系统课程的课内作业提交主页。在这里你可以提交你的代码或最终数据。

- 本课程共有三次作业。这是第一次作业，你每天可以提交一次结果。
- 截至今日起是2022-4-3, 24:00。
- 我们的数据集100%是公开数据集，因此你榜上的最好成绩就是最终成绩。

ECNU recommendation system undergraduate course competition

This is the homework (competition) home page of the recommendation system course offered by Mr. Zhang Wei. Here you can submit your code or final data individually or in teams.

- There are three assignments for this course.
- For each assignment, you can submit the results once a day. The final result will be based on your best result.

ECNU recommender systems class

This is a competition of RS class for undergraduates in ECNU.

14 days to go

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [My Submissions](#) [Submit Predictions](#)

Overview

Description

Evaluation

本次竞赛的评价指标选取的是均方根误差 [Root MSE-Score](#). RMSE的一般形式如下:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}$$

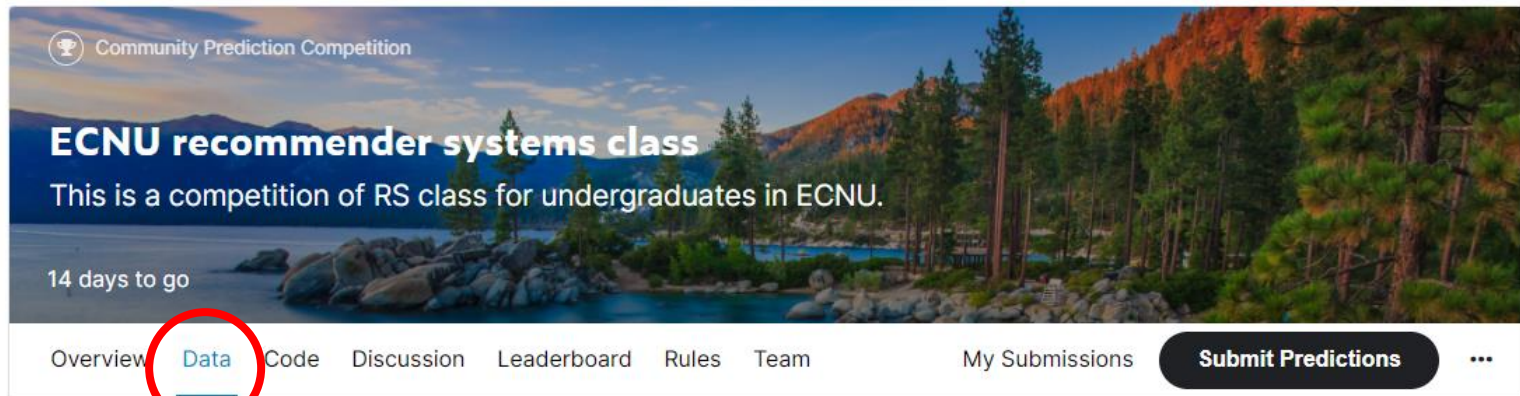
Submission Format

对于每一份提交的csv文件, 需要包含两列数据, 即主键idx和预测的stars。前者是根据test.csv里面的顺序递增产生的, 请严格按照该顺序 (采用sample_code里面的函数即可)。

文件格式形如:

```
idx,overall
0,1
1,5
2,3
.....
```

Kaggle



Data Description

数据集介绍

本次作业的数据集共分为三个文件，train.csv、test.csv和sample.csv。

数据集截取自推荐系统著名的公开数据集Amazon Dataset。如果有同学有兴趣作进一步探索可以参考[here](#)


Files

- train.csv - the training set, 即训练集
- test.csv - the test set, 即测试集
- sample_submission.csv - 你的提交样例，共两列数据：idx和overall。

Columns

- reviewerID - 用户的id。
- asin - 物品id。

Kaggle

 Community Prediction Competition

ECNU recommender systems class


This is a competition of RS class for undergraduates in ECNU.

14 days to go

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [New Notebook](#) [...](#)

[Filters](#)

[All](#) [Your Work](#) [Shared With You](#) [Bookmarks](#) [Hotness](#) [▼](#)

 **助教的示例代码sample_code**
Updated 3d ago
0 comments · ECNU recommender systems class [2](#) [...](#)